INDIRECT MEASUREMENTS OF ATMOSPHERIC TEMPERATURE PROFILES FROM SATELLITES: II. MATHEMATICAL ASPECTS OF THE INVERSION PROBLEM

S. TWOMEY
U.S. Naval Research Laboratory, Washington, D.C.

ABSTRACT

The difficulties and instabilities accompanying the inversion of radiance data to infer temperature structure are closely related to the high degree of interdependence existing among these nominally independent measurements.

The radiance measurements discussed in the accompanying papers are shown to be interdependent to a marked degree. It is shown that one of the measurements can be predicted from the others with an accuracy which is only a little worse than the experimental accuracy.

The application of this kind of analysis to determine optimum choices of measurements and the information content thereof is outlined.

1. INTRODUCTION

A review of the problems associated with indirect inferences such as the determination of temperature structure from measurements of atmospheric radiance has been given elsewhere by the writer (Twomey [3]). There it was shown that the integral equation

$$\int K(x, y)f(x)dx=g(y)$$

or the matrix relationship (to which the functional relationship is in practice sooner or later reduced)

$$\mathbf{Af}=\mathbf{g}$$

were in general unstable, and that this instability was worse the smoother the kernel. In this situation a kernel which is well behaved in the usual sense is not desirable; the optical transmittances which provide the kernels in the present problem are of course smooth and well behaved (indeed monotonic) so a considerable degree of instability must be expected.

Closely associated with the inherent instability of the inversion is the degree of interdependence among the kernels and thereby amongst the solutions. This question was discussed in the reference cited above and is most clearly exemplified in a recent paper by Mateer [2] who showed that the Umkehr data used to infer ozone distributions could be described to well within the experimental accuracy by combinations of four basic independent patterns. Hence even when 40 or 50 or more Umkehr data points are measured, only four *independent* pieces of information result. Umkehr observations give integral transform data for an exponential kernel exp $(-yx)$.

It is apparent that since infrared transmittance curves (such as those discussed in the preceding paper of this series) are essentially combinations of exponentials, one must expect a very similar result to that reported by Mateer. One cannot in this case use sets of observed values of the transform $g(y)$ as a basis, as Mateer was able to do, since the data do not yet exist. However, the possible degree of independence can be inferred by examining the kernels themselves. In view of the cost of adding data points to a satellite experiment, it is obviously important to make such an analysis.

2. THE INTERDEPENDENCE OF THE MEASUREMENTS

If, say, six measurements

$$g_i=\int_a^b K_i(x)f(x)dx+\epsilon_i \qquad (i=1, 2, \ldots, 6)$$

are made in the presence of some error "noise", it may be possible to find a combination of five kernels which so nearly approximate the sixth that the sixth measurement can be predicted in advance to better than the experimental error. It is also possible that two or more measurements can be thus predicted from the remainder.

To determine the degree of independence among N measurements g_i such that

$$g_i=\int_a^b K_i(x)f(x)dx+\epsilon_i$$

where the ϵ_i are a set of randomly distributed errors with r.m.s. value e, say, the above equation is multiplied by an arbitrary factor ξ_i and is summed over all i, to give:

$$\sum_i \xi_i g_i=\sum_i \xi_i\epsilon_i+\int_a^b \sum_i \xi_i K_i(x)f(x)dx.$$

Since the ξ_i are arbitrary, they may be normalized by a requirement that $\Sigma\xi_i^2=1$. Then

$$g_i=-\xi_i^{-1}\sum_{k\neq i}\xi_k g_k+\xi_i^{-1}\left\{\sum_k \xi_k\epsilon_k+\int_a^b \sum_k \xi_k K_k(x)f(x)dx\right\}.$$

This equation provides a formula for predicting one of

the g_i from measurements of the others; the first term on the right side is completely dependent on the other g's, and represents the predictable part; the bracketed expression is unpredictable, i.e., independent of the other measurements—it is therefore the only part of g_i containing new information. When the two terms in this expression are equal, the error and the information-bearing parts are equal, and one may speak of this as a "signal-to-noise" ratio of unity. The process of finding the best prediction formula is that of varying the ξ_i so as to make $\left|\int_a^b \Sigma \xi_i K_i(x) f(x) dx\right|$ as small as possible for all functions $f(x)$ which may be encountered. If thereby a set of ξ_i is obtained such that

$$|\sum \xi_i \epsilon_i| \geq \left|\int_a^b \sum_i \xi_i K_i(x) f(x) dx\right| \tag{1}$$

for all f's which are likely to occur, then one of the g_i can be predicted, with an uncertainty of prediction which is no larger than the observational error. The expected value of $|\Sigma \xi_i \epsilon_i|$ is merely the r.m.s. error e for random errors. For any continuous and, hence, bounded $f(x)$, it is possible to write

$$\int_a^b K_i(x) f(x) dx = C \int_a^b K_i(x) dx$$

where C is a finite number and $f(x)$ takes the value C at least once in (a, b); hence it is possible to find a number f_M such that

$$\left|\int_a^b \sum \xi_i K_i(x) f(x) dx\right| = \left|f_M\right| \left|\int_a^b \sum \xi_i K_i(x) dx\right|.$$

Hence (1) will hold if

$$\left|\int_a^b \sum \xi_i K_i(x) dx\right| \leq |e/f_M|$$

and the question of mutual independence can be examined by considering the minimal value which $\left|\int_a^b \sum_i \xi_i K_i(x) dx\right|$ can be made to attain by a suitable selection of the ξ_i. By the Schwartz inequality

$$\left|\int_a^b \sum \xi_i K_i(x) dx\right| \leq \sqrt{b-a} \sqrt{\int_a^b [\sum \xi_i K_i(x)]^2 dx}$$

$$= \sqrt{b-a} Q^{1/2}.$$

Q is the quadratic form whose associated matrix is

$$\left|\left|\int_a^b K_i(x) K_j(x) dx\right|\right|;$$

its minimum value is the smallest λ_1 of the eigenvalues of the matrix, and it attains this minimum when $(\xi_1, \xi_2, \ldots, \xi_N)$ coincides with the corresponding eigenvector (Courant-Hilbert [1] (pp. 31–34)). The smallest value attained by $\left|\int_a^b \sum \xi_i K_i(x) dx\right|$ is therefore less than $\sqrt{b-a}\sqrt{\lambda_1}$. If this is less than $|e/f_M|$, i.e., if $\sqrt{\lambda_1} \leq |e f_M^{-1}|/\sqrt{b-a}$, at least one of the g_i can be predicted from the remainder with a precision comparable to the experimental error; if n of the eigenvalues are less than $e^2 f_M^{-2}/(b-a)$, n of the g_i can be so predicted and no more than $N-n$ independent pieces of information can be derived.

In the problem of temperature sounding by infrared measurements, the kernels $K_i(x)$ can represent either the transmittance functions or their derivatives, depending on whether the radiative transfer equation is written (see equation (9) of Wark and Fleming in this series)

$$g(\nu) = -\int_0^{x_0} B \frac{d\tau}{dx} dx$$

or integrated by parts to give

$$g(\nu) = -[\tau B]_0^{x_0} + \int_0^{x_0} \tau \frac{dB}{dx} dx.$$

Because of the numerical problems of differentiation, it is better here to operate with the second form. The variable x can be any variable which decreases monotonically through the atmosphere. In the following the average of the set of transmittances was used as the variable—it has the necessary properties, and the additional convenience of ranging from 0 to 1 as well as the merit of physical meaning. However, it is obvious that the choice of variable is not critical. The variable used here will be designated x.

The magnitude of $e/|f_M| \cdot \sqrt{b-a}$ must be estimated for comparison with the eigenvalues. e is of the order of 0.5 erg/(cm.^{2} sec. strdn. cm.$^{-1}$); for the present x, $\sqrt{b-a}$ is unity; f_M is now the probable maximum of dB/dx; while an exact figure is not needed it is important for a realistic appraisal that the value chosen be not too large (which would lead to an over-optimistic appraisal of the degree of independence, one which would only be valid were very large variations in atmospheric temperature to be expected) or too small (which would lead to an over-pessimistic assessment). It is important to realize that the measure of independence is meaningful only as applied to the manifold of functions from which actual indicial functions may be drawn; if arbitrarily large excursions of temperature and lapse rate occurred from time to time at all levels in the atmosphere, then any N distinct measurements would be independent.

For atmospheres such as actually occur, a fair estimate for $|dB/dx|_{max}$ is several hundred, in the appropriate units. (It should be noted that this quantity is not a local maximum of the derivative, but a number sufficient to make the inequality

$$\left|\int_0^1 \tau_i(x) \frac{dB}{dx} dx\right| \leq \left|\frac{dB}{dx}\right|_{max} \left|\int_0^1 \tau_i(x) dx\right|$$

true for each i and every $B(x)$ which may occur. It is therefore, by the mean-value theorem, an upper bound of

an average in the large, and is little affected by large gradients of limited extent; it is in reality closer to the mean value of B than to the local maxima of dB/dx.) In terms of the eigenvalues of the matrix

$$\|\int_0^1 \tau_i(x)\tau_j(x)dx\|$$

therefore, N independent pieces of information cannot be derived from the data with a signal-to-noise ratio of S or better if any of the N eigenvalues satisfies the inequality (in round numbers)

$$\sqrt{\lambda} \leq 10^{-3}S.$$

It should however be noted that the inequality is sufficient, but not necessary. Necessary conditions do not appear to be obtainable, but the whole question can be avoided if the minimal eigenvector is used to produce the linear combination of kernels $\Sigma\xi_i\tau_i(x)$. The smallness of the resulting function and its inner products with representative indicial functions can then be used directly and compared with the errors.

3. APPLICATION TO EXPERIMENTAL DATA

The application of these procedures to experimental data is illuminating and may serve to dispose of the impression (which some recent papers convey) that the problems and limitations of indirect sensing spring from the methods of solution rather than from the fundamental nature of what is measured. The four transmittance functions for ground measurements by James (part IV of this series) were used to compute the matrix $\|\int \tau_i(x)\tau_j(x)dx\|$, for which the eigenvector associated with the smallest eigenvalue (0.000057) was (0.04325, −0.5552, 0.7849, −0.2719). The linear combination $0.04325\tau_1 - 0.5552\tau_2 + 0.7849\tau_3 - 0.2719\tau_4$ was therefore that with the smallest magnitude (norm), hence the best estimate which could be made of one of the measurements was that of I_3: this was evidently given by neglecting the (small) linear combination $0.04325\,I_1 - 0.5552\,I_2 + 0.7849\,I_3 - 0.2719\,I_4$. Thus $I_3 \cong -0.0551\,I_1 + 0.7074\,I_2 + 0.3464\,I_4$. Inserting actual measurements (131.01, 132.39, 133.07) for I_1, I_2, I_4 one obtains:

$$I_3 \cong 132.53.$$

The actual measurement of I_3 was:

$$I_3 = 132.85 \pm 0.5.$$

Hence that measurement was predictable to better than the experimental error, and in that sense was redundant.

For the satellite case discussed by Wark and Fleming in the preceding paper, a similar procedure was followed using their six transmittance curves. The matrix $\Psi = \|\int \tau_i(x)\tau_j(x)dx\|$ was computed and its eigenvalues and eigenvectors derived.

The results were as follows (rounded mainly to three significant figures):

$$\Psi = \begin{pmatrix} 80.8 & 89.3 & 91.4 & 92.9 & 94.0 & 94.8 \\ 89.3 & 100.0 & 102.9 & 105.0 & 106.5 & 107.5 \\ 91.4 & 102.9 & 106.7 & 109.4 & 111.4 & 112.8 \\ 92.9 & 105.0 & 109.4 & 112.7 & 115.3 & 117.0 \\ 94.0 & 106.5 & 111.4 & 115.3 & 118.4 & 120.5 \\ 94.8 & 107.5 & 112.8 & 117.0 & 120.5 & 122.9 \end{pmatrix};$$

eigenvalues, 7.18, 0.595, 0.000184, 0.0733, 0.00384, 634; eigenvectors,

$$\begin{pmatrix} .624 \\ .418 \\ .132 \\ -.128 \\ -.362 \\ -.520 \end{pmatrix} \begin{pmatrix} -.667 \\ .382 \\ .445 \\ .249 \\ -.0930 \\ -.376 \end{pmatrix} \begin{pmatrix} .00365 \\ -.0978 \\ .381 \\ -.654 \\ .602 \\ -.236 \end{pmatrix} \begin{pmatrix} .207 \\ -.646 \\ .230 \\ .522 \\ .142 \\ -.441 \end{pmatrix} \begin{pmatrix} -.0338 \\ .310 \\ -.647 \\ .212 \\ .540 \\ -.384 \end{pmatrix} \begin{pmatrix} .350 \\ .395 \\ .410 \\ .421 \\ .431 \\ .437 \end{pmatrix}.$$

The smallest eigenvalue is comparable to, but larger than, the acceptable limit; one may infer that a sixth piece of independent information *may* be represented in the data; whether it can be inferred with any degree of confidence requires a more detailed analysis.

The eigenvector provides a set of coefficients which by the previous analysis can be used in two ways

(i) to obtain an approximation of one transmittance curve in terms of the others

(ii) to obtain a prediction of one radiance from the others.

The approximation for the fourth transmittance curve is, of course, obtained by neglecting the linear combination $0.00365\tau_1 - 0.0978\tau_2 + 0.381\tau_3 - 0.654\tau_4 + 0.602\tau_5 - 0.236\tau_6$; in other words, it is $\tau_4 \approx 0.00558\tau_1 - 0.1496\tau_2 + 0.583\tau_3 + 0.921\tau_5 - 0.360\tau_6$.

This approximation gives the fourth transmittance with considerable accuracy. Table 1 shows for various levels the actual transmittance and that computed by the above formula. The actual and "synthetic" transmittances are indistinguishable at the 1 percent level from the top of the atmosphere to about $\tau \approx 0.2$; deviations up to 10 percent of the actual transmittances exist below this level (through the troposphere), but since τ is less than 0.2 in this region a 10 percent change in transmittance will have quite a small effect. The same coefficients can, of course, be applied to the radiances themselves. From the sets of radiances used by Wark and Fleming for Key West, Little

TABLE 1.—*Comparison of actual transmittance (697-cm.$^{-1}$ channel) with that given by the linear combination of the other five channel transmittances*

p(mb.)	τ(actual)	τ(linear combination)	p(mb.)	τ(actual)	τ(linear combination)
0.100	1.000	1.000	64.8	0.645	0.646
1.13	.991	.989	78.4	.583	.582
2.31	.982	.981	90.4	.531	.529
3.73	.972	.971	109	.456	.453
5.72	.959	.957	139	.354	.350
10.1	.932	.932	168	.269	.268
14.1	.909	.908	194	.208	.212
20.7	.871	.870	234	.133	.146
27.5	.833	.834	283	.070	.079
36.6	.784	.786	360	.020	.014
44.3	.744	.745	415	.007	−.011
53.6	.698	.699	552	.000	−.025

Rock, Seattle, and Churchill for the first, second, third, fifth, and sixth channels, the linear combination predicts the fourth radiances to be 39.67, 40.43, 39.19, and 32.84 erg/(cm.^{2} sec. strdn. cm.$^{-1}$), whereas they were actually 40.49, 41.33, 39.61, and 33.35 erg/(cm.^{2} sec. strdn. cm.$^{-1}$), respectively. Thus the fourth radiances are not entirely dependent on the others (as also inferred from the eigenvalues), but the error noise, shown by Wark and Fleming to be acceptable up to about 0.5 erg/(cm.^{2} sec. strdn. cm.$^{-1}$), is comparable to the informative parts of these radiances.

4. FURTHER APPLICATIONS

The functions $\tau_1(x)$, $\tau_2(x)$, . . ., $\tau_N(x)$ provide a set of skew vectors in N dimensional function space; the functions

$$\phi(x)=\xi_1\tau_1(x)+\xi_2\tau_2(x)+\ldots+\xi_N\tau_N(x),$$

when the ξ_i are elements of the eigenvectors of Ψ taken in turn, provide an orthogonal set of function-vectors. The latter are not, however, normal—their norms equal the corresponding eigenvalues. This is readily demonstrated: let $\xi=(\xi_1, \xi_2, \ldots, \xi_N)$ and $\zeta=(\zeta_1, \zeta_2, \ldots, \zeta_N)$ be eigenvectors of Ψ with eigenvalues λ and β respectively. Then

$$\int[\xi_1\tau_1(x)+\ldots+\xi_N\tau_N(x)]\,[\zeta_1\tau_1(x)+\ldots+\zeta_N\tau_N(x)]dx=\xi^*\Psi\zeta=\beta\xi^*\zeta=\beta\xi\cdot\zeta=0$$

unless ξ and ζ are identical, in which event the value is $\lambda\xi\cdot\xi=\lambda$; the symbol * denotes the transpose.

Thus when the analysis of the previous section was carried out, an orthogonal system of functions was implicitly produced. The previous results show that one (or more) of these "directions" are but very weakly represented in the observations; i.e., the observed functions are all almost orthogonal to one or more of these function-vectors.

If a measurement is to be eliminated with the least loss of information, that $\tau_i(x)$ clearly should be chosen which is nearest to the most weakly represented base function-vector. If the orthogonal set is

$$\boldsymbol{\phi}=(\phi_1(x), \phi_2(x), \ldots, \phi_N(x))$$

and if $(u_{1j}, u_{2j}, \ldots, u_{Nj})$, $j=1, 2, \ldots, N$ are the eigenvectors of Ψ, then the sets of functions $\boldsymbol{\tau}=(\tau_1(x), \tau_2(x), \ldots, \tau_N(x))$ and $\boldsymbol{\phi}$ are related by

$$\boldsymbol{\phi}=\mathbf{U}^*\boldsymbol{\tau}$$
$$\boldsymbol{\tau}=\mathbf{U}\boldsymbol{\phi}$$

for $\mathbf{U}=\|u_{ij}\|$ is an orthogonal matrix since Ψ is a real symmetric matrix.

Any $\tau_i(x)$ can therefore be resolved into a linear combination $\Sigma_k u_{ik}\phi_k(x)$; the relative magnitude of the lth component is therefore u_{il}. The $\tau_i(x)$ "nearest" to $\phi_l(x)$ is that for which the component with respect to $\phi_l(x)$ is relatively largest. The measurement to be rejected for least loss of information is therefore that giving rise to the largest u_{il}—which follows by inspection of the eigenvalues and eigenvectors. (l is the index of the smallest eigenvalue, and is usually 1; the corresponding eigenvector will then be the first column of $\mathbf{U}$; the ith measurement is then that selected for deletion, if the ith element of the first column is the largest in the column.)

If a large number N of possible channels and a given experimental accuracy are considered, then the optimum location for any M measurements ($M<N$) can be found as follows:

(i) a complete transmittance curve is computed for each channel under consideration.

(ii) the matrix $\Psi=\|\int\tau_i(x)\tau_j(x)dx\|$ is computed and eigenvectors are found.

(iii) the number of eigenvalues larger than the permissible lower bound (e.g. the error noise level or some number of times larger) gives the number of independent pieces of information $\nu(N)$ contained in the measurement.

(iv) assume the smallest eigenvalue is the first, then the eigenvector matrix $\mathbf{U}$ is examined to find k such that u_{kl} is the largest of the set $\{u_{il}\}$, $i=1, \ldots, N$.

(v) the kth row and column are deleted from Ψ to give a new Ψ of order $N-1$.

(vi) the entire procedure is repeated for Ψ of order $M=N-1$, then $N-2$, and so on. At each stage the number of independent pieces of information $\nu(M)$ and the optimum choice of channels is obtained.

The application of the results of the analysis is obvious. A plot of ν (M) versus M, for instance, will show the return (in terms of information) versus number of channels or cost. The effect of varying instrumental resolution or accuracy can be gauged by carrying out the computations with transmissions for varying slit widths or accuracy. Many of these questions can be examined qualitatively in other ways, but an objective numerical procedure has obvious advantages.

Finally it may be remarked that a deliberate redundancy for control purposes can be readily achieved by including a measurement which is predictable from the others with a high degree of confidence. The selection of the most suitable spectral interval for this purpose can be made by the methods just described, the combinations with small eigenvalues being now sought rather than rejected.

REFERENCES

1. R. Courant and D. Hilbert, *Methods of Mathematical Physics*, vol. I, Interscience Publishers, New York, 1953.
2. C. Mateer, "On the Information Content of Umkehr Observations," *Journal of the Atmospheric Sciences*, vol. 22, No. 4, July 1965, pp. 370–381.
3. S. Twomey, "The Application of Numerical Filtering to the Solution of Integral Equations Encountered in Indirect Sensing Measurements," *Journal of the Franklin Institute*, vol. 279, No. 2, 1965, pp. 95–109.

[*Received December 8, 1965; revised April 18, 1966*]